

A STUDY ON EDUCATIONAL DATA MINING

Dr.S.Umamaheswari
Associate Professor

K.S.Divyaa
Research Scholar,

School of IT and Science
Dr.G.R.Damodaran College of Science
Tamilnadu

ABSTRACT

Data mining is called as knowledge discovery in database (KDD), is known for its powerful role in discovering hidden information from large volumes of data. Data Mining is applied in various fields and widely used in educational field to discover hidden patterns from student's data. Educational Data Mining (EDM) is a growing field, concerned with developing methods for recognising the unique characters of data that come from educational surroundings, and applying those methods to better understand students, and helps in decision making. The aim of this paper is to discuss about the educational data mining and its component, goals and method.

Keywords: EDM Components, EDM tools and objectives

1. INTRODUCTION

EDM develops methods and applies techniques from machine learning, statistics and data mining to analyse data collected during teaching and learning [1]. Educational Data Mining focuses on developing new tools and algorithms for discovering data patterns. EDM is emerging as a research area with a collection of computational and psychological methods and research approaches for understanding how students learn. Educational data mining generally

emphasizes reducing learning into small components that can be analysed and then influenced by software that adapts to the student [13]. Student learning data collected by learning systems are being researched to produce predictive models by applying educational data mining methods that separate data or find relationships. These models play a key role in building adaptive learning systems in which adaptations or interventions based on the model's predictions can be used to change what students experience next or even to recommend outside academic services to support their learning [14]. A unique feature of data mining is that they are hierarchical. Some features of EDM are session, time and content. Time is significant to capture the data. The content is important for explaining results and knowing where model can work. The EDM component, goal and methods are important to build a model.

2. EDUCATIONAL DATA MINING

The key components of EDM are Stakeholders of Education, Data mining Methods-Tools and Techniques, Educational data, Educational Environment which meet the Educational objectives (see fig. 1). Each facet in this fig.1 will be some characteristics and helpful to build or acquire a model.

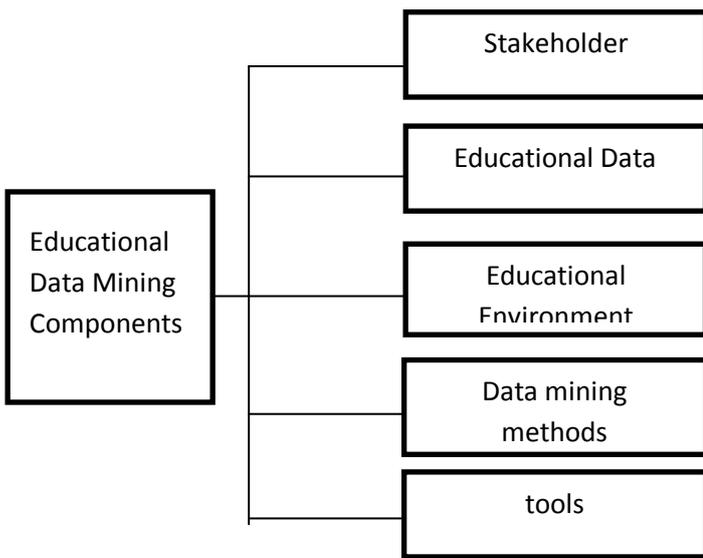


fig: 1

• **STAKE HOLDERS**

The stakeholder of the education can be divided into three classes.

Primary group: This group is directly involved with teaching and learning process. Example: Students (learners) and Faculties (teachers / learners, educators etc.)

Secondary group: This group is indirectly involved in the development of the foundation. Example: Alumni and Parents.

Hybrid group: This group required with administrative and decision making process. Example: Employers, Administrator and Educational Planner, and Experts.

• **EDUCATIONAL DATA**

Decision-making in the study of academic development involves extensive analysis of immense masses of educational information. Data are produced from heterogeneous sources like divers and variegated Uses in, diverse and distributed, structured and

unstructured data. The data are commonly generated from the offline or online source.

Offline Data: Offline Data are produced from traditional and modern classroom interaction

Interactive teaching/learning environments, learner/educators information, students' attendance, Emotional data, course information, data collected from the academic department of an institution[2].

Online Data: Online Data are created from the geographically separated stakeholder of the Training, distance educations used in, web based instruction used in, computer supported collaborative learning used in social networking sites and online group meetingspace.

Example: Web logs, E-mail, Spreadsheets, Tran scripted Telephonic Conversations, Medical records, Legal Information, Corporate contracts, Text information, publication databases Etc.

Uncertain Data: Data are generated from scientific and measurement techniques

Heterogeneity in designing Data Warehouse, sensor generated data, privacy

Conservation process data, summarization of data.

• **EDUCATIONAL DATABASE**

Formal Environment: Direct interaction with primary group stakeholder of education. Example: Face to face class room interaction.

Informal Environment: Indirect interaction with primary group

stakeholder of education. Example: web based learning, online supported tasks.

Computer Supported Environment (individual and interaction). Direct and /or Indirect interaction with all the three groups stakeholder of education.

• **DATAMINING METHODS**

DM methods are individual of the main components in EDM. It can be generally separated into two groups:

Verification Oriented is constructed on Traditional Statistics that are Hypothesis test, Analysis of Variance, etc.

Discovery Oriented are constructed on Prediction and Description-Classification, Clustering, Prediction, Relationship Mining, Neural Network, Web mining and so forth.

• **DATAMINING TOOLS**

Data mining tools that are interactive, visual, understandable, well-performing and work directly on the data warehouse/mart of the constitution could be used by front line workers. Data mining tools help customers analyze data by executing a series of actions and returning results that provide visibility into behaviors surrounding the dimensions. Data mining tools help customers analyze data by executing a series of actions and returning results that provide visibility into behavior surroundings. Each algorithm works differently to produce an output of results. In all cases, the algorithms are "trained" by exposing them to the customers' existing data sets. The

training set might include sets such as order history, payable/receivables, web navigation logs, or customer demographic information. Referable to the speed increase of educational data, there is a need to summarize the tools according to their use/features, mixed techniques and operating programs. The some of the EDM tools are MSSQL Server, Oracle Data Mining, SPSS Clementine, Intelligent Miner (IBM), CART, C4.5, ORANGE, WEKA, MATLAB, Etc.,

3. THE MAIN GOAL OF EDUCATIONAL DATA MINING

- Predicting students' future learning behaviour by creating student models that contain detailed information as students' knowledge, motivation, meta cognition, and positions;
- Discovering or educating domain models that characterize the content to be learned and optimal instructional sequences;
- Studying the effects of different kinds of academic support that can be provided by learning software;
- Evolving scientific knowledge about learning and learners through building computational models that combine models of the student, the domain, and the software's pedagogy.

To accomplish these goal there are some techniques to follow.

4. EDM objective can be classified in the following way:

a. PREDICTION

Predicting students' future learning behaviour by creating student models that incorporate such detailed information as students' knowledge, motivation, meta cognition, and attitudes. Prediction entails developing a single view of the data (predicted variable) from some combination of other views of the data (predictor variables). Predictive models have been utilized for understanding what behaviours in a learning environment and participation in discussion forums, taking practice tests and it will predict which students might fail a grade. Prediction shows promise in developing domain models, such as connecting procedures or facts with the specific sequence and the amount of practice points that best teach them, and forecasting and understanding student educational outcomes, such as success on protests after tutoring.

CLASSIFICATION

The estimation of categorization is to pose an object into one class or category, established on its other characteristics. In teaching, teachers and instructors are all the time classifying their students about their knowledge, motivation, and behaviour. Assessing exam answers is also a classification work, where a sign is set according to certain evaluation criteria.

REGRESSION

The prediction of real-valued output is regression. The things want to predict will be a numerical in Educational data mining. The most classic form of regression is a linear regression. Linear Regression performs quite good in many cases despite being too simple and particularly when you hold a great deal of

information. Regression and classification play vital role than the density estimation.

b. CLUSTERING

Clustering is to find points that naturally group together, splitting, and full data set into a set of clusters. Some of the example application are grouping students based on their learning difficulties and communication patterns, such as how and how much they use tools in a learning management system, and grouping users for purposes of recommending actions and resources to similar patterns. Data as learning resources, student perceptive interviews, and postings, in discussion forums can be analysed using techniques for working with unstructured data to press out the characteristics of the information and then bunching up the results. Clustering can be applied in any area that involves classifying, even to fix how much collaboration user's exhibit based on postings in the discussion forums.

c. RELATIONSHIP MINING

Relationship Mining is discovering relationships between variables in a data set with many variables.

ASSOCIATION RULE MINING

Association rule mining can be used for finding student mistakes that co-occur, associating content with user cases to establish recommendations for content that are likely to be interesting, or for producing changes to teaching approaches. These techniques can be applied to associate student activity, in a learning management system or discussion forums, With student grades or to investigate such

questions as why students' use of practice tests decreases over a semester of work.

SEQUENTIAL PATTERN MINING

Sequential pattern mining builds rules that capture the connections between occurrences of sequential events for example, finding temporal sequences, such as student mistakes followed by help seeking. This could be given to detect cases, such as students returning to making errors in mechanics when they are writing with more complex and critical thinking techniques. Key educational applications of relationship mining include discovery of associations between student performance and course sequences and discovering which pedagogical strategies lead to more effective or robust learning. This later area is called teaching analytics, is of growing importance and is meant to help researchers build automated systems that model how effective teachers work by mining their use of educational organizations.

d. DISTILLATION OF DATA FOR HUMAN JUDGEMENT

Distillation for human discernment is a technique that involves depicting data in a manner that enables a human to quickly place or classify features of the data. This arena of educational data mining improves machine-learning models because humans can identify patterns in, or features of, student learning activities, student behaviour's, or data involving collaboration among scholars. This approach is similar to visual data analytics.

e. DISCOVERY WITH MODEL

Discovery with models is a technique that requires using a validated model of a phenomenon it is developed through prediction, clustering, or knowledge engineering as a component in further analysis. A sample student activity distinguished from the data was map probing. An example of map probing then was hired within a second model of reading strategies and helped researchers study how the strategy varied across different experimental states. Discovery with models supports discovery of relationships between student behaviour's and student characteristics or contextual variables, analysis of research questions across a broad collection of contexts, and integration of psychometric modelling frameworks into machine-learned models.

4. CONCLUSION

In this study, the data mining methods have contributed for the development of student models. The contribution of student modelling is coming from classification methods, regression methods, clustering methods and methods for the distillation of data for human judgment. Classification, clustering and regression methods will supported to develop a validated models of a variety of complex constructs that have been embedded into increasingly sophisticated student models. Classification and regression models have afforded accurate and validated models of a broader range of student behaviour. Distillation of data for human judgment has itself facilitated the development of models of this nature, speeding the process of labelling data with reference to differences in student

behaviours, in turn speeding the process of creating classification and regression models. Finally this paper discussed about the Educational data mining components, goals and methods and tools in Educational Data Mining.

REFERENCES

- [1] Amershi, S., and Conati, C., (2009) "Combining unsupervised and supervised classification to build user models for exploratory learning environments" *Journal of Educational Data Mining*. Vol.1, No.1, pp. 18-71.
- [2] Baker, R. S. J. D. "Data mining for education." *International encyclopedia of education 7* (2010): 112-118.
- [3] Baker, R.S.J.D., and Yacef, K.(2009), "The state of Educational Data Mining in 2009:A review and future vision" *Journal of Educational Data Mining*, Vol.1,No. 1,pp.3-17.
- [4] Baker, R. S. J. d., S. M. Gowda, and A. T. Corbett. 2011. "Automatically Detecting a Student's Preparation for Future Learning: Help Use Is Key." In *Proceedings of the 4th International Conference on Educational Data Mining*, edited by M. Pechenizkiy, T. Calders, C.Conati,S.Ventura, C. Romero, and J. Stamper, 179–188.
- [5] Baker, R. S. J. d. 2011. "Data Mining for Education." In *International Encyclopedia of Education*, 3rd ed., edited by B. McGaw, P. Peterson, and E. Baker. Oxford, UK: Elsevier.
- [6] Baker, R. S. J. d., A.T. Corbett, and V. Aleven. 2008. "More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing." In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*. Berlin, Heidelberg: Springer-Verlag, 406–415.
- [7] Baradwaj,B.K., and Pal,S.(2011), "Mining Student Data to Analyze Students' Performance".*International Journal of advanced Computer Science and applications*. 2,6
- [8] Nelson,B., Nugent, R., Rupp, A. A.(2012), "On Instructional Utility, Statistical Methodology, and the Added Value of ECD: Lessons Learned from the Special Issue", *Journal of Educational Data Mining*. Vol.4, No.1, pp.224-230.
- [9] Romero, C., and Ventura S.(2013), " Data Mining in Education". *WIREs Data Mining and Know.Dis.*,Vol.3,pp.12-27.
- [10] Romero, C., and Ventura, S. (2007), " Educational Data Mining : A survey from 1995 to 2005" *Expert Systems with Applications*. Vol. 33, pp.135-146.
- [11] Tchounikine, P., Rummel, N., McLaren, M. (2010), "Computer Supported Collaborative Learning and Intelligent Tutoring Systems", *Advances in Intelligent Tutoring Systems*, SCI No.308,pp.447-463.
- [12] Romero, C., Ventura, S., Espejo, P.G., Hervas, C. (2008) *Data Mining*

Algorithms to Classify Students.
Proceedings of the First International Conference on Educational Data Mining, 8-17.

[13] Siemens, G., and R. S. J. d. Baker. 2012. "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration." In *Proceedings of LAK12: 2nd International Conference on Learning Analytics & Knowledge*, New York, NY: Association for Computing Machinery, 252–254.

[14] Marie Bienkowski ,Mingyu Feng, Barbara Means, "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief" Center for Technology in Learning SRI International, SRI International, October 2012.